

高等学校におけるデータサイエンス教育の カリキュラム開発と実践，評価基準の作成

林 宏樹¹ 笹嶋宗彦² 大里隆也³

概要：本論文は、全校生を対象とした兵庫県立姫路西高等学校のデータサイエンス教育の3年間のカリキュラム開発についての概要である。教科「数学」「情報」「課題研究」「総合的な探究の時間」における横断的なカリキュラムマネジメントにより、データサイエンスを基盤とした研究活動を実施するカリキュラム開発を行う。全校生対象の高等学校におけるデータサイエンス教育実践の1つの事例である。

また、「データサイエンス探究・研究」という学校設定科目の客観的な評価基準を作成する。

キーワード：データサイエンス教育，データ分析，情報，数学，課題研究，理数探究，カリキュラムマネジメント

Curriculum development of data science education in high school

HIROKI HAYASHI†1 MUNEHICO SASAJIMA†2
TAKAYA OHZATO†3

1. はじめに

日本の文部科学省では、国際的に活躍し得る科学技術人材等の育成のために、理数系教育に関する教育課程等に関する研究開発を行う高等学校をスーパーサイエンスハイスクール（略称 SSH）に指定している。2020年、データサイエンス教育を学校全体に実装するカリキュラム開発のモデルケースとして、兵庫県立姫路西高等学校（以後、姫路西）が採択された。データサイエンス教育のカリキュラム開発は、兵庫県立大学と帝国データバンクとの共同開発である。

2. 背景と目的

未来社会においては、広範囲かつ目的に整合したデータの取得を前提として、それらのデータに内在する本質的構造を見極め、数理的思考に基づいて解析・問題解決を行う能力、データサイエンスを活用して新たな価値を生み出し、有用なシステム構築につなげる能力が求められる。[1]

ビッグデータに溢れ、AIが発達する時代において、AIにコントロールされない人間になるためには、探究力、自己分析力、発信力を身につける必要がある。そこで本校では、これらの力を磨きあげることで、人間にしかもつこと

のできない創造力が育成されると仮定し、データサイエンス教育を通して、それらの能力が身につくカリキュラムを産学連携によって開発する。そして、3年間の実践により検証を行う。

姫路西のSSHでは、「知」の育成プログラム・「知」の検証プログラム・「知」の連携プログラムを実施することで、研究開発の題目である「高度な「知」を有するグローバルサイエンティストの育成」を目指している。本論文は、高等学校における全校生徒を対象としたデータサイエンス教育のカリキュラム開発と実践事例である3つのプロジェクトの中の「知」の育成プロジェクトの報告である。

3. データサイエンス教育を実現するためのカリキュラム

(1) 3年間のカリキュラムの概要

姫路西におけるデータサイエンス教育を実践する「知」の育成プロジェクトでは、1年次を「探究準備期間」、2年次を「探究実践期間」、3年次を「探究展開期間」と位置づけている。

全体像は以下のとおりである。

1年次「探究準備期間」において、以下の①～④のデータサイエンス教育の基礎・基本に関する授業実践を行う。

- ① データリテラシーや研究スキルを学ぶ。
- ② データ収集する方法を学ぶ。

1 兵庫県立姫路西高等学校
2 兵庫県立大学
3 (株)帝国データバンク

	第一学年	第二学年	第三学年
国際理学科	4 単位	3 単位	2 単位
普通科	2 単位	2 単位	1 単位

- ③ データ分析手法を学ぶ.
- ④ 創造性に関するメカニズムを学ぶ.
2年次「探究実践期間」において、以下の⑤によりデータサイエンスを基盤とした研究実践を行う.
- ⑤ 自分の興味関心のあるテーマの研究を行う.
3年次「探究展開期間」において、以下の⑥により研究発表等の発信を行い、⑦により後輩の研究を指導助言することで、自らのデータサイエンスの学びを振り返り、知識・技能の定着を認知していく.
- ⑥ ⑤の論文を作成し、発表を行う.
- ⑦ 研究を行う後輩の指導を行う.

(2) カリキュラムマネジメント

姫路西では、理数に関する学科である「国際理学科」と「普通科」がある。共通教科「理数探究基礎」や「理数探究」の学習過程（探究の過程）を考慮し、3(1)を実装するために、下記の通り、教育課程を編成している。[2]

(ア) 国際理学科 1年【探究準備期】

データサイエンス教育の基礎・基本を学ぶ学校設定科目「データサイエンス研究(4単位)」を実施し、それによって「社会と情報(1単位)」「課題研究(1単位)」「総合的な探究の時間(2単位)」の代替とする。

(イ) 普通科 1年【探究準備期】

データサイエンス教育の基礎・基本を学ぶ学校設定科目「データサイエンス探究(2単位)」を実施し、それによって「社会と情報(1単位)」「総合的な探究の時間(1単位)」の代替とする。

(ウ) 国際理学科 2年【探究実践期】

データサイエンスを基盤とした研究実践を行う学校設定科目「データリサーチ研究(3単位)」を実施し、それによって「社会と情報(1単位)」「課題研究(1単位)」「総合的な探究の時間(1単位)」の代替とする。

(エ) 普通科 2年【探究実践期】

データサイエンスを基盤とした研究実践を行う学校設定科目「データリサーチ探究(2単位)」を実施し、それによって「社会と情報(1単位)」「総合的な探究の時間(1単位)」の代替とする。

(オ) 国際理学科 3年【探究展開期】

データサイエンスの学びを振り返るための学校設定科目「グローバル研究(2単位)」を実施し、それによって「コミュニケーション英語Ⅲ(1単位)」「課題研究(1単位)」の代替とする。

(カ) 普通科 3年【探究展開期】

データサイエンスの学びを振り返るための学校設定科目「グローバル探究(1単位)」を実施し、それによって「総合的な探究の時間(1単位)」の代替とする。

4. データサイエンス教育

4-1 データサイエンス教育とは

文部科学省(2009)『高等学校学習指導要領 解説 数学編』における第1節数学Iの3(4)データの分析の内容と比べ、文部科学省(2018)『高等学校学習指導要領 解説 理数編』における第1節数学Iの3(4)データの分析においては、統計的探究プロセスが追加されている。[3][4]

また、IDSSPにおけるInternational Data Science in School ProjectではThe basic cycle of learning from dataが示されている。[5]

この統計的探究プロセスの5つの段階からなる「問題(Problem)－計画(Plan)－データ(Data)－分析(Analysis)－結論(Conclusion)」に基づき、姫路西で実施するデータサイエンス教育では、図1のような5つのプロセスとして細分化・具体化したデータサイエンスサイクル(DSサイクル)と名付けて授業の内容を組み立てている。

図1のサイクルの具体的な5つのプロセスは、

- ・問題－計画のPを組み合わせた「課題の定式化」(略称PP)
- ・データを細分化した「データ収集」(略称D)
- ・データを細分化した「データ研磨」(略称D)
- ・分析と同一である「データ解析」(略称A)
- ・結論の箇所を「結果の考察」(略称C)

と定め、1つ1つのプロセスを習得するように授業を展開し、このサイクルをスパイラルに実践していくことをデータサイエンス教育と定義している。

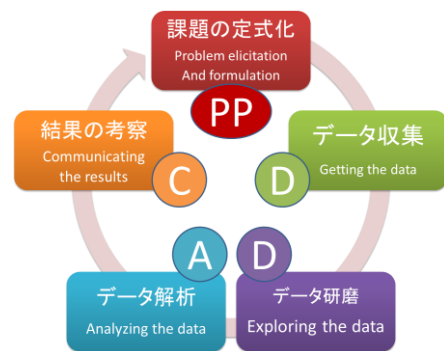


図1 姫路西高校 DS サイクル

Fig 1. Data Science Cycle of himeji Nishi high school

4-2 5つのプロセスの具体的取組

4-2-1 課題の定式化

生徒が研究テーマとなる課題を発見するために、現象を細分化して捉えたり、言葉の定義を定めることが重要である。課題の定式化のレベルが低ければデータ収集のレベル

は低くなる。

評価の要素は、

- ・言葉の定義が定まっているか
- ・課題が細分化されているか
- ・先行研究に基づいた仮説を立てているか

で判断する。

4-2-2 データ収集

e-Stat や地域経済分析システム (RESAS) などのオープンデータでのデータ収集力を養う。また、csv データを MS-Excel データの変換等、情報機器活用力も養う。

評価の要素は、

- ・必要なデータが収集できているのか
- ・自分自身でデータ採取しているのか

で判断する。

4-2-3 データ研磨

収集したデータを研磨することで、有効なデータ解析につながる。データの可視化、外れ値・異常値に関する見方や考え方を学ぶ。また、情報活用力として、MS-Excel でのデータ研磨を習得し、プログラミング言語 python を用いたデータ研磨によって、プログラミングの有用性を体験し、プログラミング言語の習得を目指す。

評価の要素は、

- ・収集したデータを自分でエクセル等を用いて加工したデータであるか

で判断する。

4-2-4 データ解析

可視化されたデータの読み取り方やデータの種類に適した可視化により結論を導く力を養う。また、相関係数や標準偏差の理解と活用、回帰分析法の理解と活用し、重回帰分析法や検定などの分析法を学ぶ。

評価の要素は、

- ・グラフの特徴的な部分に関する説明がされているか
- ・内容に適した分析方法を活用できているか

で判断する。

4-2-5 結果の考察

データ解析により、複数のデータ解析を組み合わせた結論を導いていく。最終的にはデータから読み取った内容だけで判断するのではなく、実態を調査し、データ解析の結論は1つの指標として、実態に即した課題解決する力を養う。

評価の要素は、

- ・複数の結果と結果を融合させて新たな課題や仮説を立てているか
- ・分野の異なる結果と結果を統合して新たな結論を導き出せているか

で判断する。

4.3 データサイエンスの観点における評定のつけ方

姫路西は前期・後期制であるため、前期評価をつける必

要がある。しかしながら、4-1 のサイクルを総合してデータサイエンス教育と考えているため、実際には1つ1つのプロセスごとに学んでいくので、データサイエンスの観点としての評価点を途中段階 (前期) でつける必要があるのか協議した。結果として、前期の評定をつけず、年度末において、研究を総合的に判断し、以下の積み上げ式評価5段階と加算点評価によってデータサイエンスの観点に対するチェックリストを設定した。

まず、積み上げ式評価として、

段階1 D

- ・データをグラフで可視化しているか

段階2 PP

- ・言葉の定義が定まっているか
- ・課題が細分化されているか

段階3 D

- 必要なデータが収集できているか

段階4 A

- グラフの特徴的な部分に関する説明がされているか

段階1 ができていると評価2, できていなければ評価1 段階2 ができていると評価3, できていなければ評価2 という積み上げ式で評価点をつける。

次に、加算点評価として、

PP 先行研究に基づいた仮説を立てているか

D 自分自身でデータ採取しているか

D 収集したデータを自分で MS-Excel 等を用いて加工しているか

D 内容に適した分析手法を活用できているか

C 複数の結果と結果を融合させて新たな課題や仮説を立てているか

C 分野の異なる結果と結果を統合し、新たな結論を導き出せているか

とした PPDAC に即したプロセスごとの高度な要素を含めると評価点が1上がるとして、合計 11 点満点で評価点をつけ、その評価点に基づいて5段階評定をつける。このスタイルでの評価点をつける方法によって、データサイエンスの観点の評価を異なる教員が評価点をつけても散らばりが小さくなると仮定している。

5. データサイエンス教育の実践事例

2020 年 11 月現在、1 年生国際理学科データサイエンス研究 (4 単位) を受講している生徒は、データサイエンス教育サイクルの2回目に進み、「データ解析発表会」を実施した。1 年生普通科データサイエンス探究 (2 単位) を受講している生徒は、データサイエンス教育サイクルの1回目を終えたところである。

5-1 データサイエンス教育の実践授業内容

2020 年 11 月時点での「データサイエンス研究」の授業における DS サイクルの知識・理解を習得する時間数は、

課題の定式化 2 時間，データ収集 1 時間，データ研磨 2 時間，データ解析 8 時間，結果の考察 2 時間であった。ただし，この時間数以外に研究実践としてこれらを活用する時間を設けている。以下，DS サイクルの 1 つ 1 つの詳細な内容を記す。

5-1-1 「課題の定式化」に関する授業内容

- ・生徒の興味関心に応じたテーマから理想と現実を言語化し，そのギャップから問題を発見させる。
- ・グループ討議により問題から課題を発見し，データ収集を実施させる。

5-1-2 「データ収集」に関する授業内容

e-stat 政府統計の総合窓口[6]・教育用標準データセット (SSDSE) [7]，RESAS (地域経済分析システム) [8]等からのデータ収集方法を習得させる。

5-1-3 「データ研磨」に関する授業内容

- ・欠損値や外れ値を理解し，適切な処理をさせる。
- ・ピボットテーブルを作成する技能を習得させる。
- ・散布図の作成，回帰直線の描画，決定係数の表示について習得させる。
- ・分析ツールの活用による統計量・度数分布表・ヒストグラムや箱ひげ図の作成等の可視化の技能を習得させる。

5-1-4 「データ解析」に関する授業内容

- ・平均値，標準偏差，相関係数等の数値の意味を理解し，データに適した数値を活用させる。
- ・可視化したグラフからデータの特徴を考察させる。具体例として，図 2 の体力測定データの箱ひげ図から種目間の特徴を考察させた。

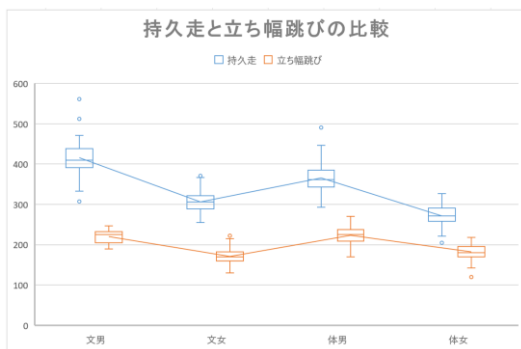


図 2 体力測定による箱ひげ図

Fig. 2 Box plot by physical fitness test

- ・回帰について，単回帰分析を理解し活用させる。
- ・重回帰分析を理解し，モデル検証を実施させる。具体例として，体力測定の結果を用いて，「ハンドボール投げ」を目的変数とし，重回帰分析によるモデル検証を理解させる。

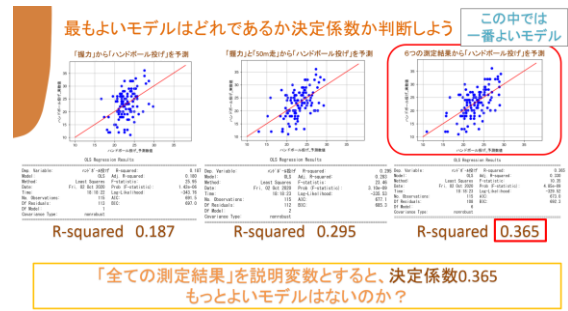


図 3 モデル検証の授業スライド

Figure 3 Model validation class slides

5-1-5 「結果の考察」に関する授業内容

- ・統計データ分析コンペティション[7]の過去の受賞作品を事例に結果の考察を実施させる。
- ・生徒の興味関心に応じたデータサイエンスに係るコンペティションの過去の受賞作品についてグループ討議させる。

5-2 データサイエンスを基盤とした研究の評価の実践

5-2-1 「データ解析発表会」の実施と評価結果

2020 年 11 月 19 日，データサイエンス研究における「データ解析発表会」を実施した。1 つの研究に対して，4 名の教員が 4-3 のデータサイエンスのチェックリスト (DS チェックリスト) を基に下記の表の通り，5 段階評価をつけた。

評価点 1	データをグラフで可視化できていない
評価点 2	データをグラフで可視化できている
評価点 3	・言葉の定義が定まっている ・課題が細分化されている
評価点 4	必要なデータが収集できている
評価点 5	グラフの特徴的な部分に関する説明がされている

結果，評価点 1 は 0%，評価点 2 は 40%，評価点 3 は 23%，評価点 4 は 30%，評価点 5 は 8%であった。また，教員 4 人が 1 グループごとに 5 段階評価した結果，10 グループの標準偏差を比較すると，2 グループが 0.5，3 グループが 0.58 に対して，他の 5 グループは 1 前後であった。

また，「データ解析発表会」では「課題発見力」「課題解決力」「発信力」の 3 つの観点のルーブリックによる 5 段階評価も同時に実施した。3 つの観点と DS チェックリストの評価点のそれぞれの相関を調べると，「課題解決力」との相関係数が 0.94 となり，強い正の相関が見受けられた。

5-2-2 評価に関する考察

今回の「データ解析発表会」で初めて DS チェックリストを活用した評価を実施した。

教員による評価の結果を考察すると，5-1-1 の通り，5 グループは概ね散らばり小さい評価となったが，5 グループは標準偏差が 1 前後という大きい散らばりであった。この

事実と、「課題解決力」と強い正の相関があることは、関係していると考えている。生徒の研究指導をしていく上で、最もハードルが高いのは課題の定式化にある「課題の細分化」である。本校生徒の社会情勢の認識不足や語彙不足等、現実的な実態を深く考察したことのない生徒が多く、世の中の出来事を自分で深く洞察する習慣が乏しく、インターネットのニュース等の情報を鵜呑みにする生徒が非常に多いと感じている。そのため、抽象的な課題設定をした生徒は、データ解析のためのデータ収集で行き詰まることが多い。逆に、課題が細分化された研究は、内容に適したデータ収集がしやすいため、データ解析結果も仮説通りになる傾向が見受けられた。

そのため、本校の DS チェックリストは評価点3が「課題が細分化されている」という基準である。そのため、全体の評価点の割合をみると、5-2-1 の通り、評価点2と評価点4の割合が高く、評価点3が低いのは上記の生徒の現状が理由であると考えている。また、4人の教員の評価が散らばっているグループの評価は、評価点3がおらず、評価点4と評価点2だけという評価であった。つまり、評価点3をクリアすると評価点4になりやすいという結果を表していると思われる。

5-3 データサイエンスを基盤とした研究の指導の展望

データサイエンスを基盤とした研究の指導において、統計的探究プロセスを踏まえ、最初の乗り越えるべき壁は、課題の細分化である。ここをクリアしなければ、課題解決のためのデータ収集が難しく、データ解析することもできないことが多い。また、データ解析の結果を考察する上でも、結果と現実の実態と照らし合わせて結論を導くことが重要である。そのため、データを扱う上で、データの裏側を見抜くための知識をもち、洞察する習慣をつけることで、課題の細分化を乗り越える指導助言を心がけていく。その指導がデータサイエンスの学びを充実させることにつながると考えている。

また、重回帰分析を含む回帰分析法や検定などの指導を実施したが、まだ研究に活用した生徒が非常に少なかった。今後、生徒が主体的に様々な分析手法や python の活用できるよう指導の工夫をする。

評価においては、今回の取組を基に、異なる教員が評価をしても散らばり小さい評価基準となるチェックリストやルーブリックの作成を目指していく。

6. おわりに

本稿では、高等学校における全校生対象のデータサイエンス教育のカリキュラムの実装と現在(2020年11月)における授業実践事例を報告した。今後は、データサイエンス教育サイクルの各プロセスでの指導事例や身についた能力を客観的に分析した結果など、本校の取り組んでいるデー

タサイエンス教育の検証を行う。また、産学連携により開発した教材や、python を利用したデータ分析の実践事例や教材開発をし、それらの効果の検証を実施する。そして、データサイエンス教育を特定の教員が指導できるカリキュラムや授業内容とするのではなく、すべての教員が実践できるよう、体系化を目指していく。

参考文献

- [1] “大学の数理・データサイエンス教育教科方策について”
https://www.mext.go.jp/b_menu/shingi/chousa/koutou/080/gaiyou/1380788.htm (参照 2020-7-28).
- [2] 長尾篤志 (2019.9.4) 『教育課程部会 長尾主任視学官発表資料「理数探究」の充実と STEAM 教育について』
- [3] 文部科学省 (2009) 『高等学校学習指導要領 解説 数学編』
- [4] 文部科学省 (2018) 『高等学校学習指導要領 解説 理数編』
- [5] “Introductory Data Science”
<http://idssp.org/files/IDSSPFrameworks1.0.pdf>
(IDSSP Curriculum Team September 2019).
- [6] e-stat 政府統計の総合窓口 <https://www.e-stat.go.jp/>
- [7] 統計データ分析コンペティション
<https://www.nstac.go.jp/statcompe/>
- [8] RESAS 地域経済分析システム
<https://resas.go.jp/#/13/13101>